DOCUMENT RESUME

ED 185 525 CS 005 391

AUTHOR Flippo, Rona F.
TITLE A Proposal: The Need for Comparison Studies of College Students' Reading Gains in Developmental Reading Programs Using General and Specific Levels of Diagnosis.
PUB DATE Feb 80
NOTE 20p.; Paper presented at the Annual Meeting of the Southeastern Regional Conference of the International Reading Association (5th, Nashville, TN, February 6-9, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; Comparative Analysis; *Developmental Programs; Diagnostic Teaching; Postsecondary Education; Program Effectiveness; Reading Achievement; *Reading Diagnosis; Reading Programs; *Reading Research; Reading Skills; *Reading Tests; Remedial Reading; *Test Validity
IDENTIFIERS *Nelson Denny Reading Test; *Stanford Diagnostic Reading Test

ABSTRACT

This paper discusses an investigation that compared college students' reading gains when using general or specific diagnosis. It reports that significant gains were made on many subskills of the Stanford Diagnostic Reading Test by the group that had the prescriptions developed as a result of a differential (specific) diagnosis and that there were no significant gains noted on the Nelson-Denny Reading Test, a general survey test. An analysis of two other research studies on this topic shows that results of this study support the results of a study by G. L. Cox and J. E. Swalm and contradict the results of a study by B. D. Smith. It is noted, however, that since Smith used the Nelson-Denny Reading Test to test gains from specific diagnosis, her findings are not necessarily valid and not necessarily contradictory to those of the other two studies. The evidence in the literature indicating that the Nelson-Denny Reading Test is not an appropriate test for evaluating the reading ability of college students in developmental programs and that survey tests are often used inappropriately as measurements of achievement in college reading programs is discussed in the light of the findings from these research reports. (MKM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185525

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Rona F. Flippo

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A Proposal: The Need for Comparison Studies of College Students' Reading Gains in Developmental Reading Programs Using General and Specific Levels of Diagnosis

Rona F. Flippo

This article is a rebuttal to an article published by the Journal of Reading (Smith, 1977). The research investigation (Flippo, 1975) exploring a comparison of college students' reading gains when using differential or more specific diagnosis and college students' reading gains when using a survey or less specific diagnosis invalidates Smith's findings. Significant gains were made in many subskills by the group that had their prescriptions developed as a result of the differential diagnosis.

A thorough review of the literature from 1930 through 1978 reveals only two research studies that explored this type of comparison (Cox and Swalm, 1973; Smith, 1974). In her 1977 article Smith never mentions the Cox and Swalm study. Evidently, at the time Smith (1974) completed her dissertation, from which she wrote the article in question (Smith, 1977), the National Reading Conference Yearbook, which contained the results of the Cox and Swalm (1973) study, had not been available. Although the Cox and Swalm study and the Smith study are the beginnings of a body of research in this area, they both have their limitations in the relatively small sizes of their groups, Cox and Swalm (n=98) and Smith (n=75). Additionally, the findings of the two studies are contradictory to each other; however, the validity of the Smith study has been severely limited by the research design of the study because it did not allow for the

CS005391

posttesting of the groups on all the tests with which they were pretested and because Smith, the researcher, was also the instructor of both groups.

The purpose of the Cox and Swalm (1973) study was to determine if formal diagnosis before instruction is beneficial as compared to informal diagnosis during instruction; however, because of the nature of their study, they also compared formal or differential initial diagnosis to a more general or survey-type initial diagnosis. High-risk college freshmen at Rutgers University were pretested with the McGraw-Hill Reading Test and the comprehension section of the Nelson-Denny Reading Test and randomly assigned to two groups. The students in both groups were taught by eight different instructors. Methods and course content were the same for the groups, except for the amount of diagnosis each received from test results. One group was prescribed after an analysis of specific test items on sub-areas of the McGraw-Hill, or after differential diagnosis. The other group was prescribed on the overall scores in each of the three major sub-areas of the McGraw-Hill. Both groups were posttested at the conclusion of the term with another form of the same tests. Students receiving the more differential initial diagnosis scored significantly higher at the .05 level than students in the other group on the McGraw-Hill and the comprehension section of the Nelson-Denny reading tests. The results of the Cox and Swalm study indicate that instruction based upon the diagnosis of specific reading strengths and weaknesses is superior to skill instruction based on general evaluation of students' reading abilities.

Smith (1974) examined the value of additional diagnosis in comprehension and phonics in a developmental reading program at Kennesaw Junior College. The students were pretested with the Nelson-Denny Reading Test, Form A, the Van Wagenen Reading Scales, and the California Phonics Survey, and randomly assigned to two groups, both taught by Smith. Reading instruction for one group was based on the diagnostic information gained from all three tests. Reading instruction for the other group was based on the survey results of the Nelson-Denny. The only difference between groups was the amount of reading diagnosis each received from the pretests. The Nelson-Denny Reading Test, Form B, was the only test administered to both groups as a posttest at the end of the term. An analysis of covariance showed no significant difference at the .05 level in raw score gain between the groups on the Nelson-Denny; however, the group receiving the differential diagnosis did show an observed gain higher than the group having just the survey diagnosis. The results of Smith's study, using the Nelson-Denny data, indicate that instruction based on additional diagnosis beyond the survey test is not more successful than instruction based on survey test scores alone. The findings of Smith (1974) appear to be contradictory to the findings of Cox and Swalm (1973). But a third study (Flippo, 1979) indicates that the Smith findings are not necessarily valid and not necessarily contradictory to those of Cox and Swalm.

A Third Study

A third study (Flippo, 1979) was designed to compare the effects of specific and general diagnosis on postsecondary students' progress in selected reading subskills after developmental instruction. In-

struction was prescribed from the specific or differential results of a diagnostic reading test for some students and from the general and more traditionally used results of a survey reading test for other students. The selected subskills were literal comprehension, inferential comprehension, word meaning, word parts, phonetic analysis, structural analysis, scanning and skimming, and fast reading, as measured by the Stanford Diagnostic Reading Test (SDRT), Blue Level (1976), and vocabulary and comprehension, as measured by the Nelson-Denny Reading Test (NDRT), Forms C and D (1973).

The Nelson-Denny was selected because it had been used in both the Cox and Swalm, and Smith studies and because surveys had indicated that the Nelson-Denny is the test most widely used in developmental college reading programs (Lowe and Stefurak, 1970; Goodwin, 1971; Sweiger, 1972; Landsman and Cranny, 1978). The rate subtest of the Nelson-Denny was used in writing prescriptions for the control group as is traditionally done, but not as a dependent variable in this study. Farr (1968) criticizes the rate subtest of the NDRT for lacking convergent and discriminant validity, and he suggests it should not be used at all.

The Stanford Diagnostic, Blue Level, was developed to meet the reading evaluation needs of grade 9 through junior/community college students. Van Roekel (1978), in his review of the newest editions of the SDRT, stated, "The SDRT has few peers among group diagnostic reading tests" (p. 1299). This test met the needs and suggestions indicated for college developmental reading by Ironside (1969), Tittle and Kay (1971), and Evans and Dubois (1972). It was the best diag-

nostic instrument available for this population at the time of the study.

The sample consisted of approximately 450 entering freshman students enrolled in fifteen different sections of "Effective Reading" in the 1978 fall semester at the University of South Carolina's College of General Studies, Columbia. The researcher's classes and evening classes were not included in the study. These fifteen sections were taught by nine different instructors; students were randomly assigned to an experimental (x) or control ($\underset{\sim}{x}$) group so that there was approximately an equal number of each group in each section. The freshmen were required to take "Effective Reading" because they scored 390 or below on the verbal portion of the SAT. The actual number of students included in the study from the original group of approximately 450 were the 226 students (x=112, $\underset{\sim}{x}$=114) who completed two counseling-prescription appointments, a minimum of 10 lab hours in the Developmental Center, and all pre and posttests on the SDRT and NDRT.

On the basis of information from the subtests of the diagnostic instrument (SDRT), individual prescriptions for reading skill improvement were written for the students in the experimental group. Only the information from the subtests of the survey instrument (NDRT) were used in writing prescriptions for the control group. The original diagnoses from the pretest information on the students in both groups were not altered in any way throughout remediation. All students were treated with two hours per week of classroom instruction with their instructor and one hour per week of lab instruction. Lab instruction was based

on each student's individual prescription. Students in both the control and experimental groups had the same treatment except for the amount of diagnostic or differential information available for the initial counseling and prescription of instruction.

Students were not aware of the differentiation of diagnosis or the control and experimental groups. All the students were pretested with both the diagnostic (SDRT) and survey (NDRT) instruments. Students in the experimental group had all of their tests handscored and individual profiles were made using raw scores and stanines from the SDRT and raw scores and grade equivalents from the NDRT. Students in the control group only had their NDRT scored, and individual profiles were made from only the NDRT data. The SDRT's for control students were left unscored and locked in a file cabinet until the end of the term.

The diagnosis and prescription of the two groups were accomplished by use of a card system designed by the researcher especially for the testing instruments and materials selected for this study (Flippo, 1980). The card system was programmed by the researcher cataloging all the lab materials into one or more of the ten selected subskills for this study, plus the subskill of rate from the NDRT. Graduate assistants were trained to use this card system to make appropriate decisions regarding the prescription of materials according to the pretest results when counseling students in both groups. Instructors were given the pretest results used for counseling and prescription, with the more specific diagnostic information on each of their experimental students and only the more general survey test information on their control students.

At the end of the semester, all students were posttested with different forms of the SDRT and NDRT. The control group's SDRT pretests were then scored and the raw score data recorded. The gain in the ten selected subskills was measured by the pretest and posttest differences of the two groups. A t-pool test was used at the .05 level of significance.

Findings

There was a statistically significant difference in the means of the gain scores of the experimental or differentially diagnosed students in the subskills of literal comprehension, inferential comprehension, word meaning, phonetic analysis, structural analysis, and scanning and skimming. There was not a statistically significant difference between the means of the gain scores of the experimental students and the control students in the subskills of word parts, fast reading, vocabulary, and comprehension. However, on the subskills of fast reading and vocabulary, the observed gain of the experimental group was substantially higher than that of the control group, but not significantly higher at the .05 level. (See summary table showing comparison of raw score gains on each of the ten selected subskills.)

This study's findings generally support the conclusions of Cox and Swalm (1973) that postsecondary students do better in developmental reading programs that utilize more specific or differential diagnosis and prescription than in reading programs using more general testing of reading abilities. However, in this study, as in the Smith (1974) study, the gains from the Nelson-Denny (vocabulary and comprehension) did not reflect evidence of a significant difference. If this study

had solely used the Nelson-Denny to measure the gains, as Smith did, the findings would have supported the conclusions of Smith rather than those of Cox and Swalm.

Gains made as a result of diagnostic or differential prescription may not necessarily be reflected in the results of survey reading tests. Smith's study does not allow for this possibility. If Smith would have given the diagnostic tests as posttests as well as pretests to both her groups and compared those results between groups, she might have found that the group with the differential diagnosis did better as measured by the diagnostic tests. Unfortunately, this is not known; therefore, the Smith study cannot be considered valid.

These findings suggest that the Nelson-Denny Reading Test, used more than any other in college reading programs, is not consistently a good discriminator of students that have improved more and students that have improved less in developmental programs. In this study (Flippo, 1979) the gains on the subskill of comprehension, as measured by the Nelson-Denny, do not reflect the gains made on the subskills of literal and inferential comprehension, as measured by the Stanford Diagnostic. The t values of literal and inferential comprehension indicate that there was a significant difference between the gains of the groups; yet, there was not significant or observed difference between the means of the groups on comprehension, as measured only by the Nelson-Denny. Additionally, the gains made on the subskill of vocabulary, as measured by the Nelson-Denny, do not significantly reflect the gains made on the subskills of word meaning and literal comprehension (vocabulary in context), as measured by the Stanford Diagnostic. It should be noted that both groups improved to approximately the same degree on

the su[illegible]ll of word parts. There is no clear explanation for this phenomenon, and further study is needed in this area.

Misuse of Test

There has been evidence in the literature which indicates that the Nelson-Denny might not be an appropriate test for evaluating the reading ability of college students in developmental programs and that survey tests are often used inappropriately as a measurement of achievement in college reading programs. Murphy and Davis (1949) found that scores on the Nelson-Denny could be greatly increased when students were encouraged to mark every item even if they did not know the answers to the items. This would cause an inflation of gain scores from pre to posttesting without any remediation at all. Orr (1965) found the comprehension passages of the Nelson-Denny to be long, involved, and grammatically complex. Townsend (1965) suggested that the Nelson-Denny was a challenging academic type of test useful for screening good readers going into the liberal arts but would not adequately differentiate reading skills of college students. Farr (1968) pointed out that the rate section of the Nelson-Denny lacks both convergent and discriminant validity.

Ironside (1969) and Evans and Dubois (1972) cautioned against using survey tests for the purpose of assessing achievement in reading or for making a diagnosis. Goodwin (1971) found that 60% of the 300 junior college reading teachers he surveyed considered diagnostic the survey test given at the beginning of the term as a screening device, and the Nelson-Denny was the most widely used. Chester, Dulin, and Carvell (1975) suggested that changes should be made in the comprehen-

sion questions of the Nelson-Denny. They suggested semantic and syntactic simplification of stems and roots; specificity and clarity of questions that served as a way of identifying information about how well the students understood the stimulus material being tested, not how well they could answer certain types of questions; and questions that students would actually need to ask and to answer for themselves outside of the testing situation in order to understand the reading materials.

The literature seems to indicate that the Nelson-Denny is a survey test that is suitable for screening academically able college-bound students. It does not adequately sample the abilities of vocational-education students or postsecondary students in need of remediation programs. This test is not suitable for the community college/developmental education market for screening purposes and certainly not for assessing achievement in reading in any college setting. It seems that the critical flaw of the Smith study and article is the misuse of this test. This misuse is chronic in the literature and research relative to postsecondary developmental reading programs.

Recommendations

Based on the statistical findings and educational implications of this study and the opinion of this researcher, the following recommendations are made for further research and consideration. Additional research investigating a comparison of specific and general levels of diagnosis in college reading programs is needed. The Nelson-Denny should not be used for purposes beyond that of screening developmental reading students and should be restricted for screening if there is access to a more appropriate test for this population. More diagnostic

reading tests designed specifically for measuring the reading and study skills of postsecondary students would be desirable, and dissemination of information to community college/developmental reading practitioners on the presently available tests is gravely necessary.

Summary Table
Results of t-Pool Test Comparing Raw Score Gains
$t(224, .05) = 1.960$[1]

Subskills/Test	Group	Number	Mean Gain	Standard Deviation	Standard Error	t value	Significant Difference
Literal Comprehension (SDRT)	Experimental	112	3.517	2.910	.274	4.140	yes
	Control	114	1.576	4.030	.377		
Inferential Comprehension (SDRT)	Experimental	112	2.857	2.667	.252	5.526	yes
	Control	114	.614	3.385	.317		
Word Meaning (SDRT)	Experimental	112	2.375	2.812	.265	5.594	yes
	Control	114	- .096	3.753	.351		
Word Parts (SDRT)	Experimental	112	2.357	3.221	.304	.541	no
	Control	114	2.596	3.422	.320		
Phonetic Analysis (SDRT)	Experimental	112	2.205	3.649	.344	4.394	yes
	Control	114	.070	3.654	.342		
Structural Analysis (SDRT)	Experimental	112	2.517	4.024	.380	2.124	yes
	Control	114	1.447	3.540	.331		
Scanning and Skimming (SDRT)	Experimental	112	3.473	3.193	.301	6.401	yes
	Control	114	.061	4.668	.437		

Summary Table
Results of t-Pool Test Comparing Raw Score Gains
$t(224, .05) = 1.960$[1]
Continued

Subskills/Test	Group	Number	Mean Gain	Standard Deviation	Standard Error	t value	Significant Difference
Fast Reading (SDRT)	Experimental	112	3.133	3.588	.339	1.522	no
	Control	114	2.368	3.956	.370		
Vocabulary (NDRT)	Experimental	112	4.526	3.815	.360	1.623	no
	Control	114	3.675	4.062	.380		
Comprehension (NDRT)	Experimental	112	5.883	4.699	.444	.056	no
	Control	114	5.842	6.290	.589		

[1]Bruning, J. L. and Kintz, B. L. Computational handbook of statistics. Glenview, Illinois: Scott, Foresman and Company, 1968, p. 219.

REFERENCES

Brown, G. M. & Cottrell, A. B. California Phonics Survey. Monterey, California: California Test Bureau/McGraw-Hill, 1963.

Brown, J. I., Nelson, M. J., & Denny, E. C. Nelson-Denny Reading Test, Forms A and B. Boston: Houghton Mifflin Company, 1960.

Brown, J. I., Nelson, M. J., & Denny, E. C. Nelson-Denny Reading Test, Forms C and D. Boston: Houghton Mifflin Company, 1973.

Bruning, J. L., & Kintz, B. L. Computational handbook of statistics. Glenview, Illinois: Scott, Foresman and Company, 1968.

Chester, R. D., Dulin, K. L., & Carvell, R. Mature readers' Nelson-Denny comprehension scores with and without syntactically - and semantically - simplified item - stems and/or item - foils. In G. H. McNinch and W. P. Miller (Eds.), Reading: convention and inquiry. Twenty-fourth Yearbook of the National Reading Conference. Clemson, S. C.: National Reading Conference, 1975, 227-234.

Cox, G. L., & Swalm, J. E. A formalized diagnostic approach to study skills instruction for specially-admitted college freshmen. In P. L. Nacke (Ed.), Programs and practices for college reading (Vol. 2). Twenty-second Yearbook of the National Reading Conference. Boone, N. C.: National Reading Conference, 1973, 105-110.

Evans, H. & Dubois, E. Community/junior college remedial programs - reflections. Journal of Reading, 1972, 16, 38-45.

Flippo, R. F. Comparison of college students' reading gains in a developmental reading program using general and specific levels of diagnosis. Unpublished doctoral dissertation. University of Florida, 1979.

Flippo, R. F. A Diagnostic/Prescriptive/Counseling/Cataloging System. Innovative Strategies. San Antonio, Texas: Special Interest Group for College Reading Improvement, International Reading Association, May 1980.

Farr, R. The convergent and discriminant validity of several upper level reading tests. In G. B. Schick and M. M. May (Eds.), Multidisciplinary aspects of college adult reading. Seventeenth Yearbook of the National Reading Conference. Milwaukee: National Reading Conference, 1968, 181-191.

Goodwin, D. D. Measurement and evaluation in junior college reading programs. Junior College Research Review, October 1971, 6 (2), 1-3. (ERIC Document Reproduction Service No. ED 053 714)

Ironside, R. A. Who assesses reading status and progress-tests, teachers, or students? Paper presented at the Twelfth Annual Meeting of the College Reading Association, Boston, Massachusetts, March 13-15, 1969. (ERIC Document Reproduction Service No. ED 031 374)

Karlsen, B., Madden, R. & Gardner, E. F. Stanford Diagnostic Reading Test, Blue Level, Form A and Form B. New York: Harcourt Brace Jovanovich, 1976.

Landsman, M. B., and Cranney, A. G. Training and activities of Florida community college reading teachers. Florida Reading Quarterly, 1978, 14, (2), 17-22.

Lowe, A. J., and Stefurak, D. W. The college reading improvement programs of Georgia: 1969-70. In G. B. Schick and M. M. May (Eds.), Reading: process and pedagogy (Vol. 2). Nineteenth Yearbook of the National Reading Conference. Milwaukee: National Reading

Conference, 1970, 118-124.

Murphy, H. D. & Davis, F. B. A note on the measurement of progress in remedial reading. Peabody Journal of Education, 1949, 27, 107-111.

Orr, D. B. Nelson-Denny Reading Test. In O. K. Buros (Ed.), The sixth mental measurements yearbook. Highland Park, New Jersey: The Gryphon Press, 1965, 1077-1078.

Raygor, A. L. McGraw-Hill Basic Skills System: Reading Test, Forms A and B. Monterey, California: McGraw-Hill, 1970.

Smith, B. D. The effect of differential diagnosis upon reading improvement for high risk junior college students (Doctoral dissertation, Georgia State University, 1974). Dissertation Abstracts International, 1975, 35, 4995A-4996A. (University Microfilms No. 74-23, 1978)

Smith, B. D. Do we need differential diagnosis at the college level? no. Journal of reading, 1977, 21, 62-66.

Sweiger, J. D. Designs and organizational structure of junior and community college reading programs across the country. In F. P. Greene (Ed.), College reading: problems and programs of junior and senior colleges (Vol. 2). Twenty-first Yearbook of the National Reading Conference. Boone, N. C.: National Reading Conference, 1972, 1-7.

Tittle, C. & Kay, P. Selecting tests for an open admissions population. Paper presented at the annual meeting of the American Educational Research Association, New York, February, 1971. (ERIC Document Reproduction Service No. 048 359)

Townsend, A. Nelson-Denny Reading Test. In O. K. Buros (Ed.), The sixth mental measurements yearbook. Highland Park, New Jersey: The Gryphon Press, 1965, 1078-1080.

Van Roekel, B. H. Stanford Diagnostic Reading Test. In O. K. Buros (Ed.), The eighth mental measurements yearbook (Vol. 2). Highland Park, New Jersey: The Gryphon Press, 1978, 1297-1300.

Van Wagenen, M. J. Van Wagenen Analytical Reading Scales. Minneapolis, Minnesota: M. J. Van Wagenen, 1966.